



# Collaboration a Proposed Path Forward

Deb Agarwal  
DAAgarwal@lbl.gov  
Lawrence Berkeley Laboratory



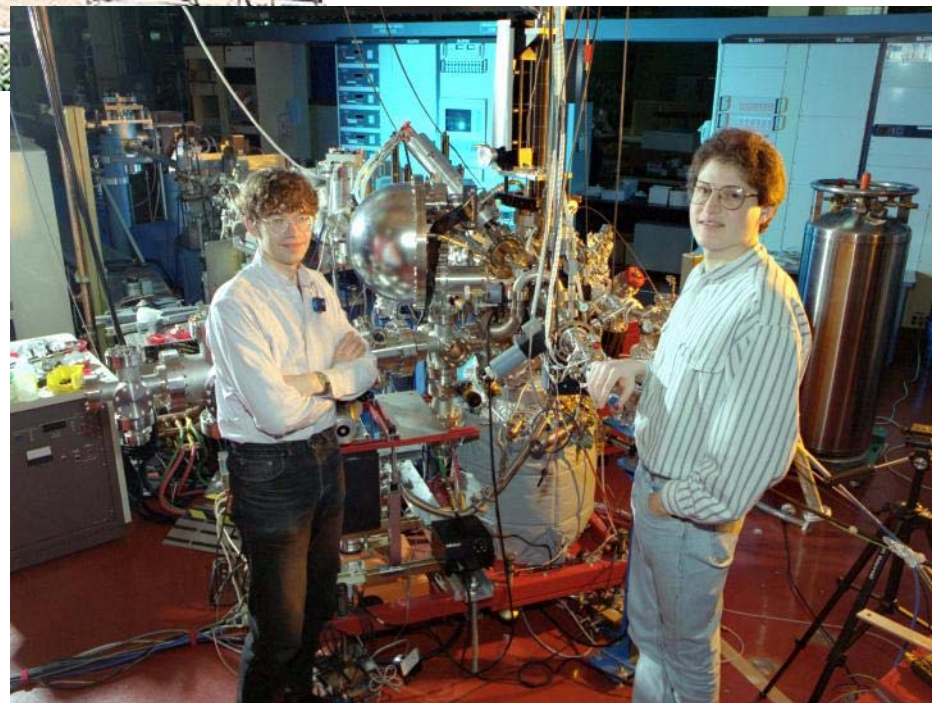
# Evolution of Collaboration



- Distributed Collaboratory Experiment Environments
  - Remote access to instruments – ALS, EM, NMR, Tokamak
  - Technology development – electronic notebooks, security, multicast, etc
- DOE 2000
  - Technology development – communication, security, shared spaces, logbooks, etc
  - Collaboratory pilots – On-line instruments (MMC) and Shared resources (DCC)
- National Collaboratories Program
  - Technology development – distributed computing, security, collaborative tools, portals, web services, data management, etc
  - Collaboratory pilots – PPDG, ESG, NFC, and CMSC
  - Partnering among agencies
- Science of Collaboratories Program



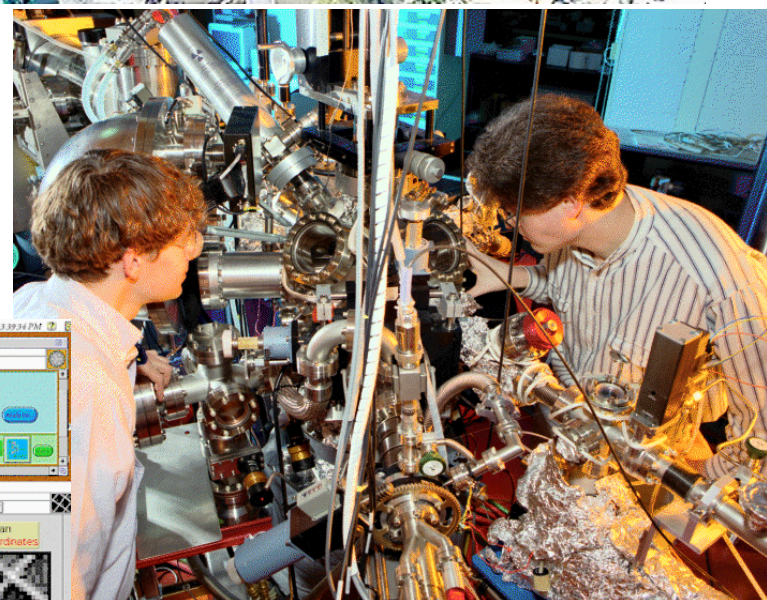
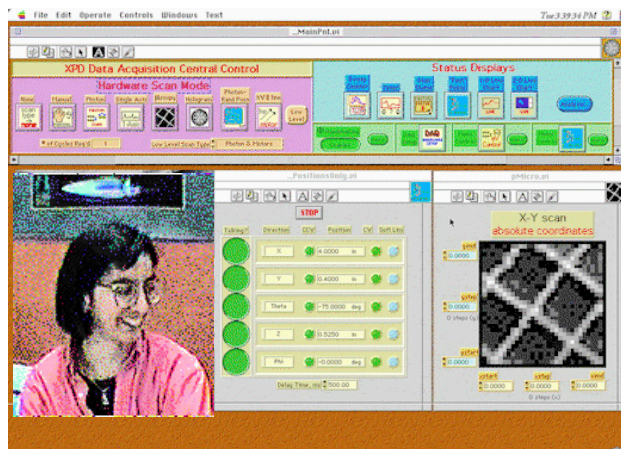
## Example - Advanced Light Source







# Spectro-Microscopy Collaboratory



Univ. of Wisconsin

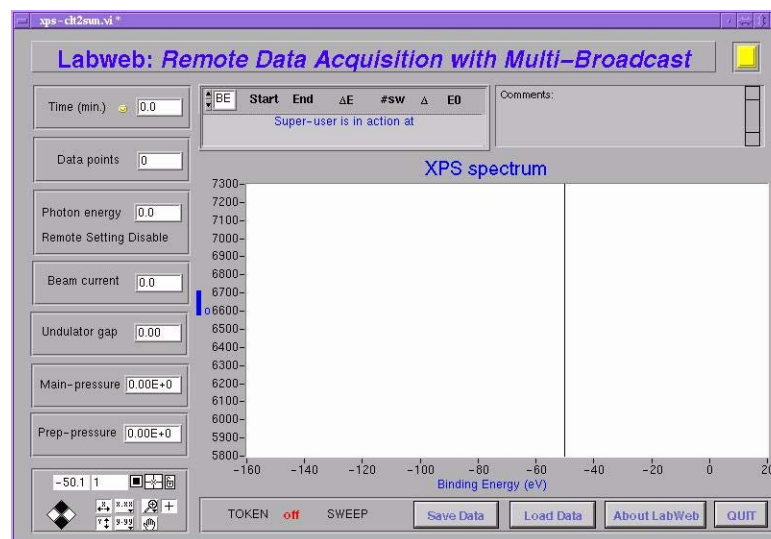
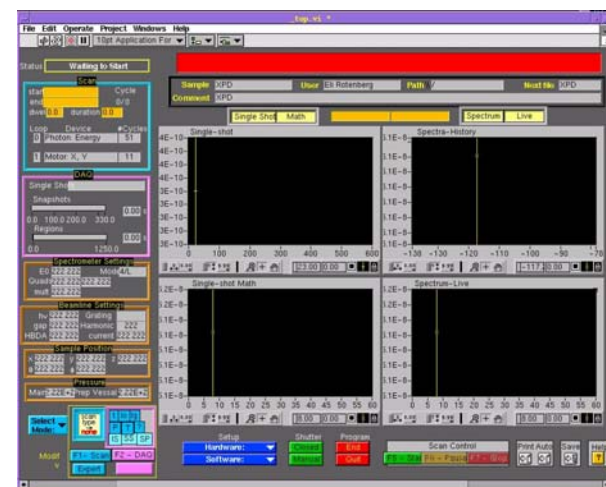
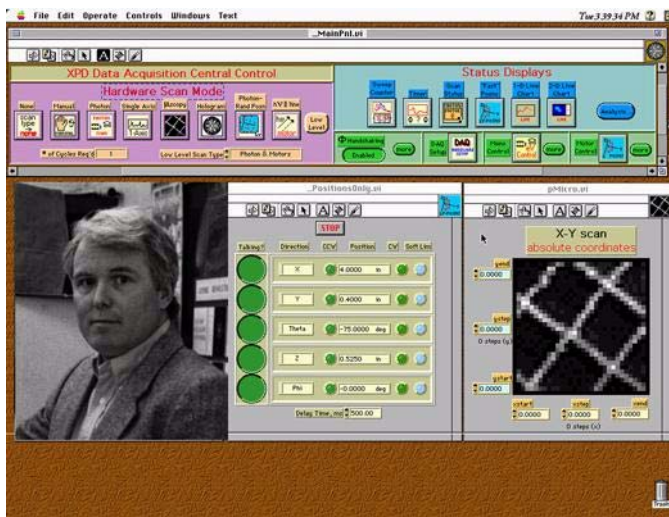
LBNL

Network





# Experiment Control Panel Evolution







# Confctlr



**ConfCntrl**

v0.4   Status   ☐ ☐   Contact Host:

☐ ☐   Remote User:

☐ Security   ☐ Settings   ☐ Conference   ☐ Quit

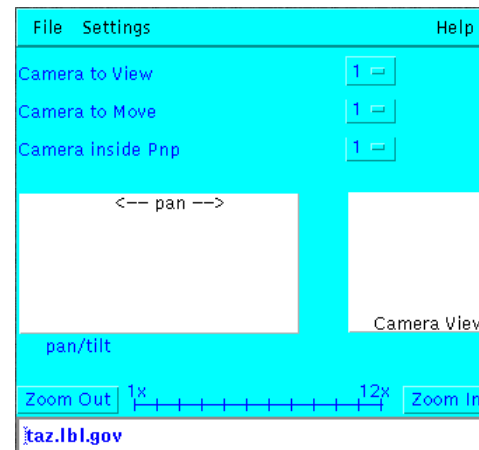
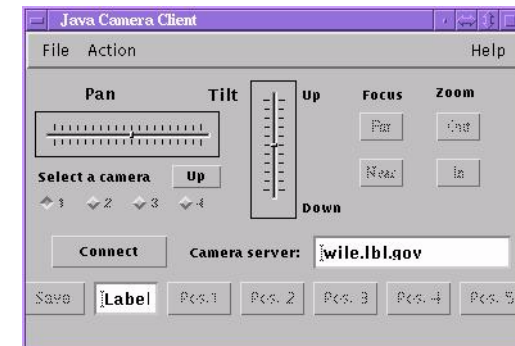
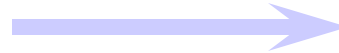
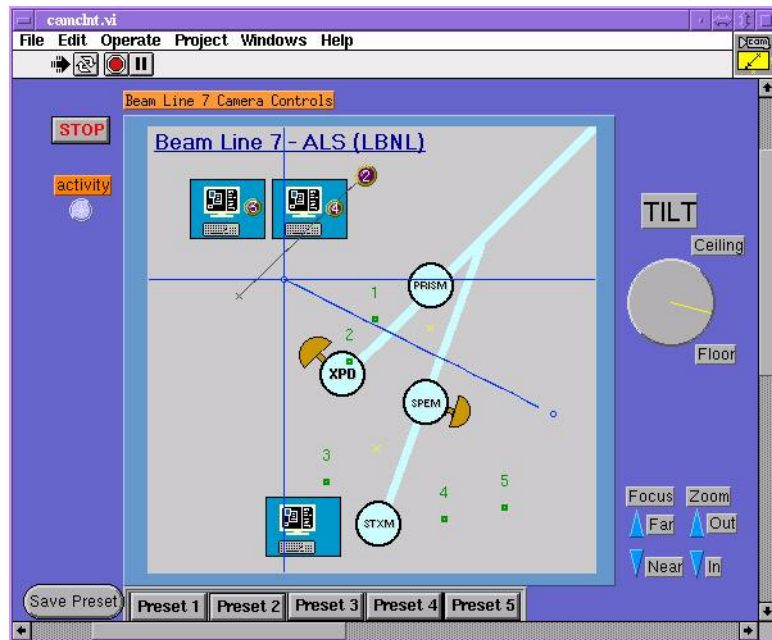
Host:    TTL:

**Conference Name**

	Address		Port		Local	Remote
video 1	<input type="text" value="224.2.195.166"/>	<input type="button" value="v"/> <input type="button" value="d"/>	<input type="text" value="30270"/>	<input type="button" value="v"/> <input type="button" value="d"/>	<input checked="" type="checkbox"/> Stop	<input checked="" type="checkbox"/> Stop
video 2	<input type="text" value="224.2.195.166"/>	<input type="button" value="v"/> <input type="button" value="d"/>	<input type="text" value="30272"/>	<input type="button" value="v"/> <input type="button" value="d"/>	<input type="checkbox"/> Start	<input checked="" type="checkbox"/> Stop
audio	<input type="text" value="224.2.195.166"/>	<input type="button" value="v"/> <input type="button" value="d"/>	<input type="text" value="30274"/>	<input type="button" value="v"/> <input type="button" value="d"/>	<input type="checkbox"/> Start	<input checked="" type="checkbox"/> Stop
					<input checked="" type="checkbox"/> Stop All	<input checked="" type="checkbox"/> Stop All



# Remote Camera Control





# Lessons Learned



- Collaboration takes effort and thus must
  - Provide a perceptible benefit to all participants
  - Fit with current work practices
  - Be accessible to the users
- Collaboration tools need to be used regularly (not on the shelf)
- Group must already have a strong need to collaborate
- Collaboration technology creates new paradigms of interaction
- Difficult to predict how a technology will be used by a particular group
- Support for asynchronous interaction important
- Sociology is a dominant factor





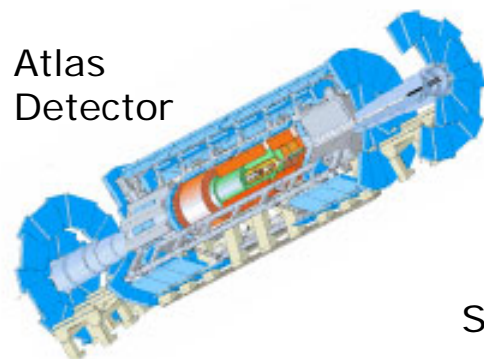
# Today's Distributed Science Reality



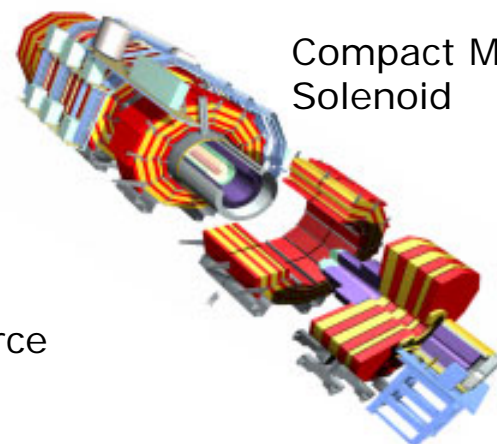
- Collaborations include as many as 1000's of scientists
- Collaborators located all over the world
- Many users never visit the site
- Virtual organization involved in managing the resources
  - Include multiple sites and countries
  - Distributed data storage
  - Distributed compute resources
  - Shared resources
- No one site or group controls the computers users are accessing resources from
- High performance computing, networking, and data transfers are core capabilities



# Experiments

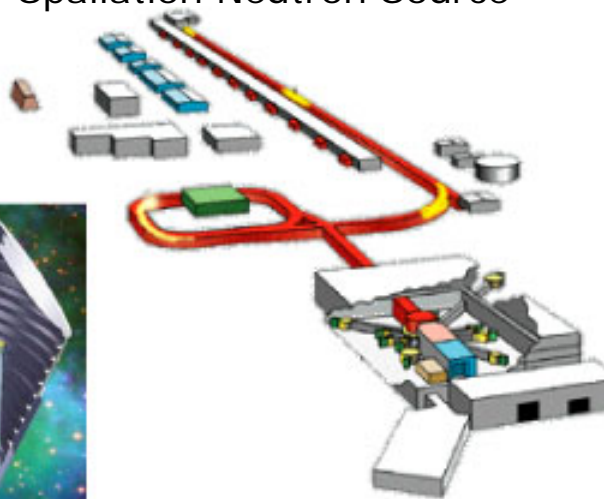


Atlas  
Detector

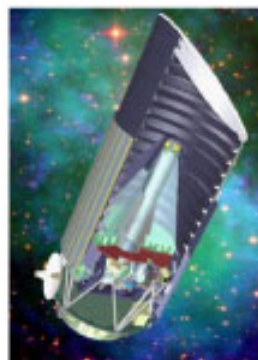


Compact Muon  
Solenoid

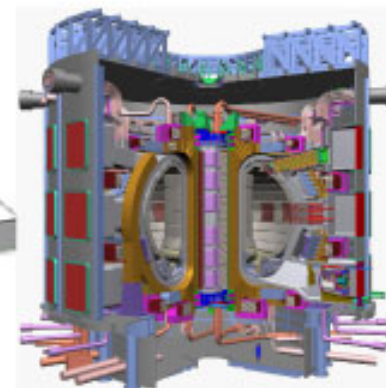
Spallation Neutron Source



Ultrahigh Voltage  
Electron Microscope



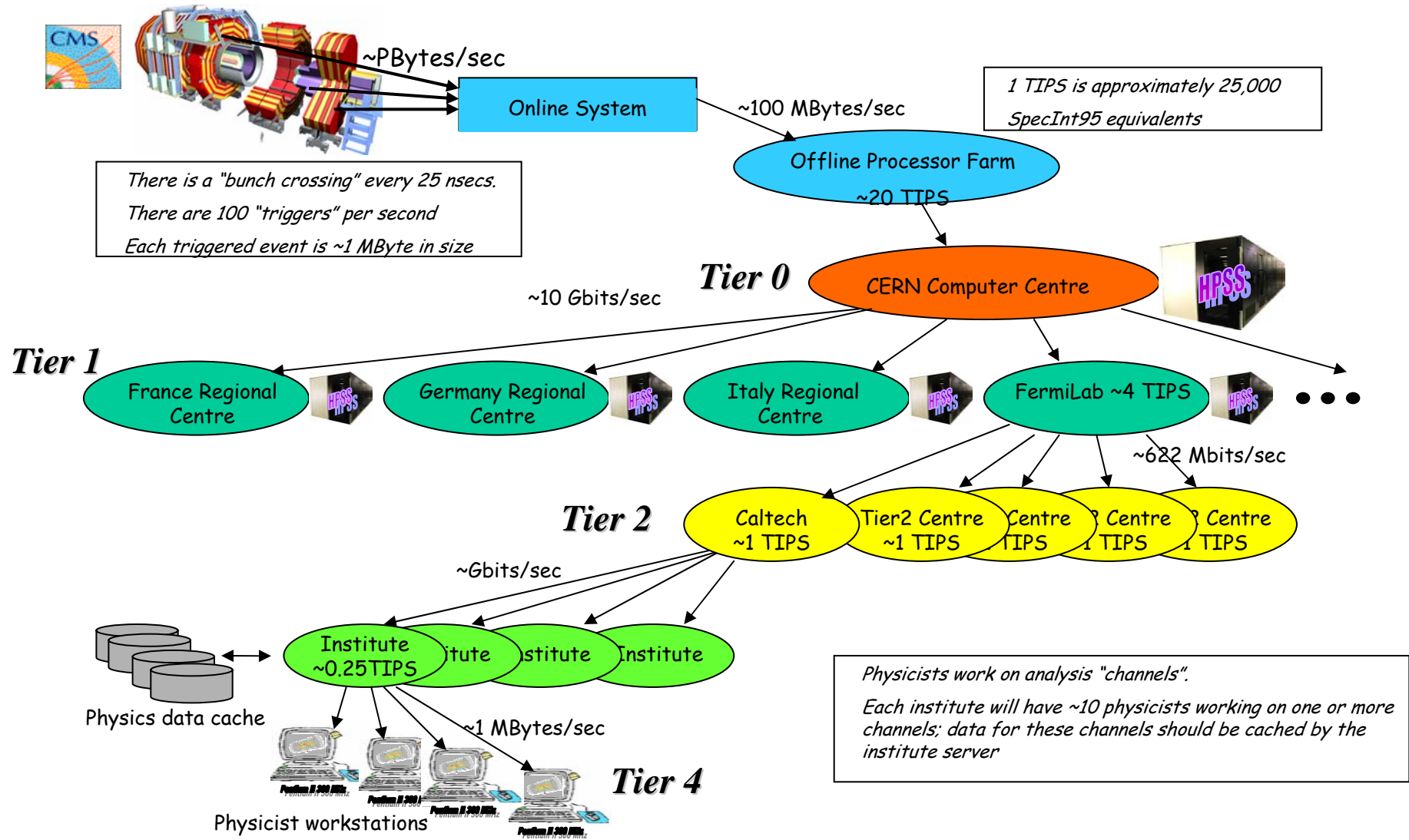
Supernova/  
Acceleration  
Probe



ITER Tokamak



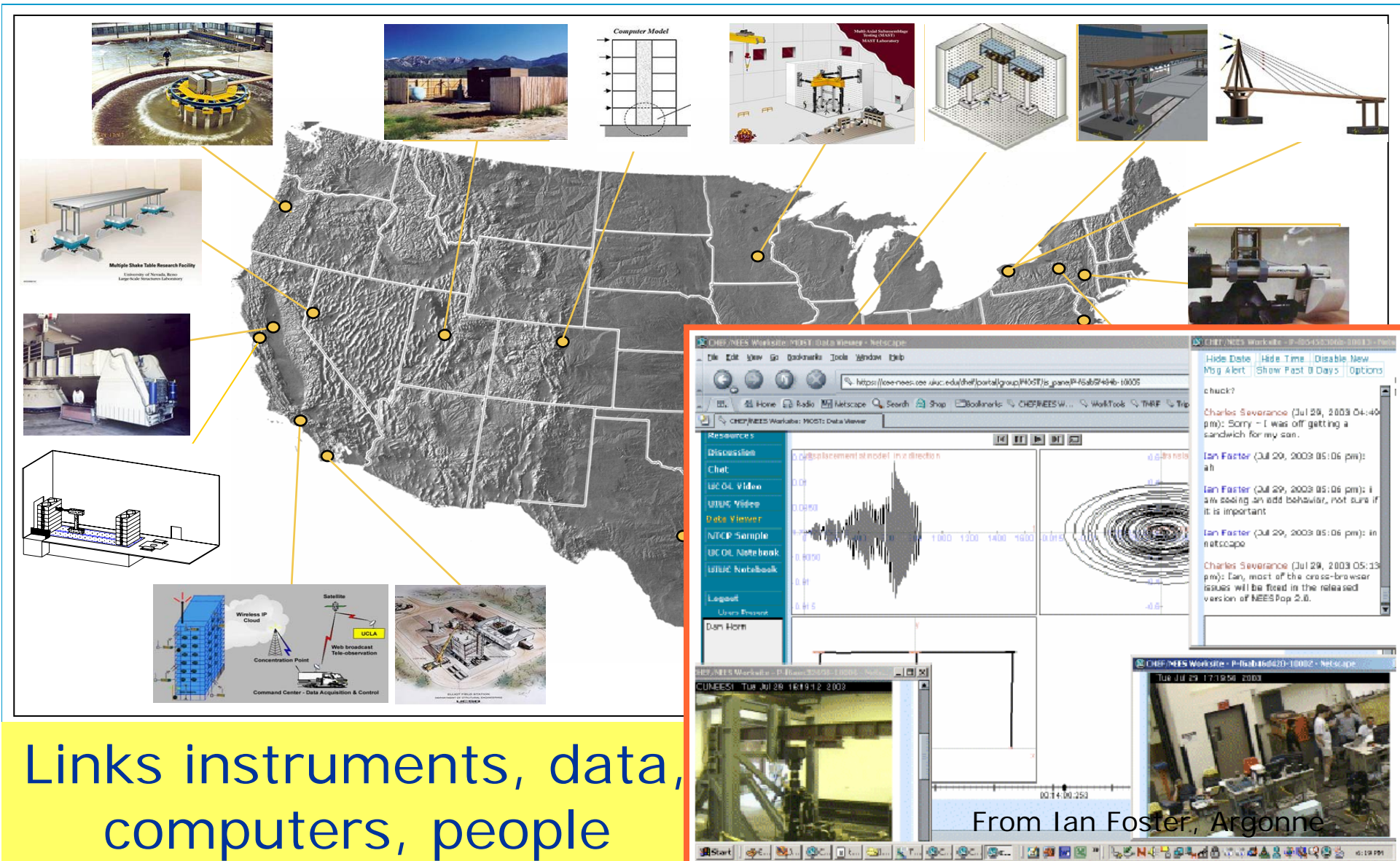
# Distributed Science Infrastructure in High Energy Physics



from Harvey Newman, CalTech



# NSF Network for Earthquake Engineering Simulation





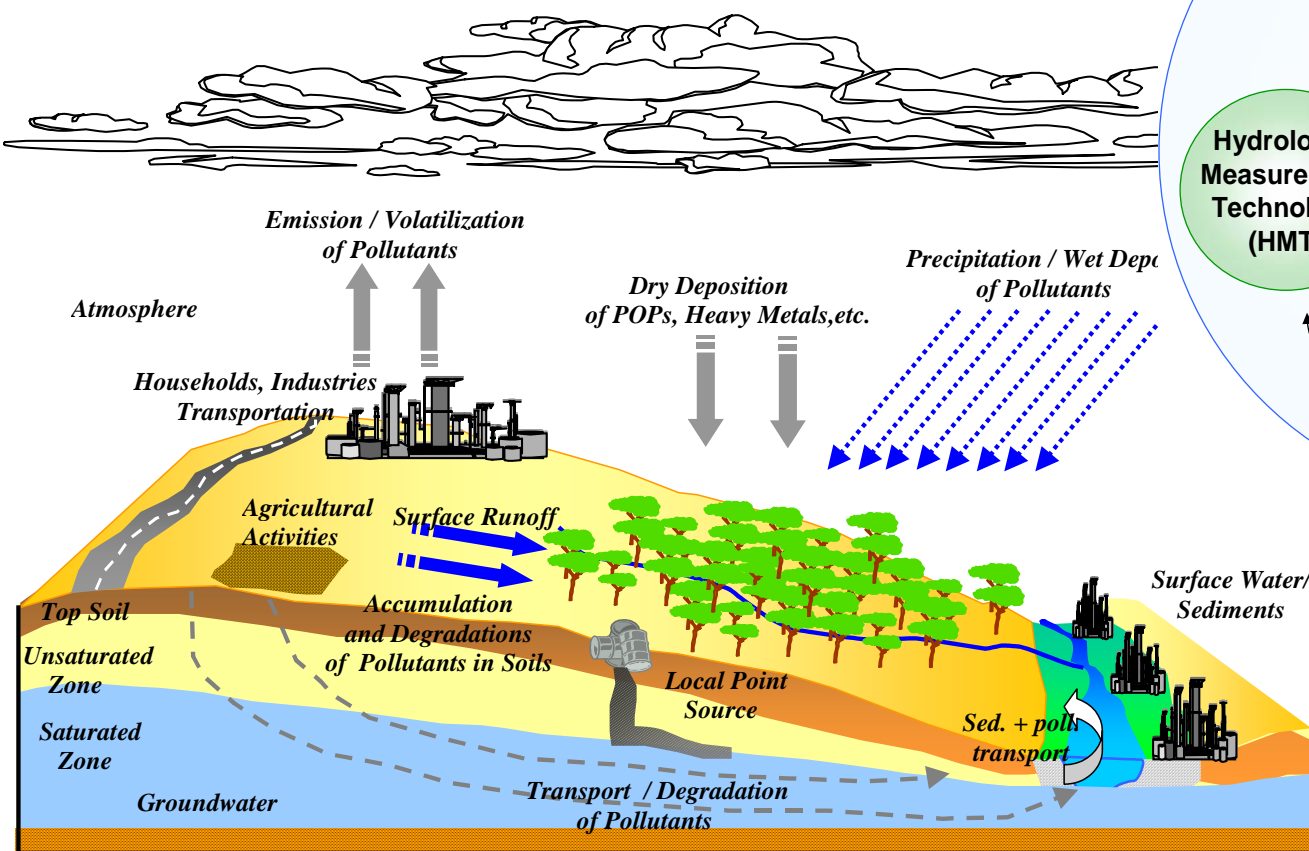
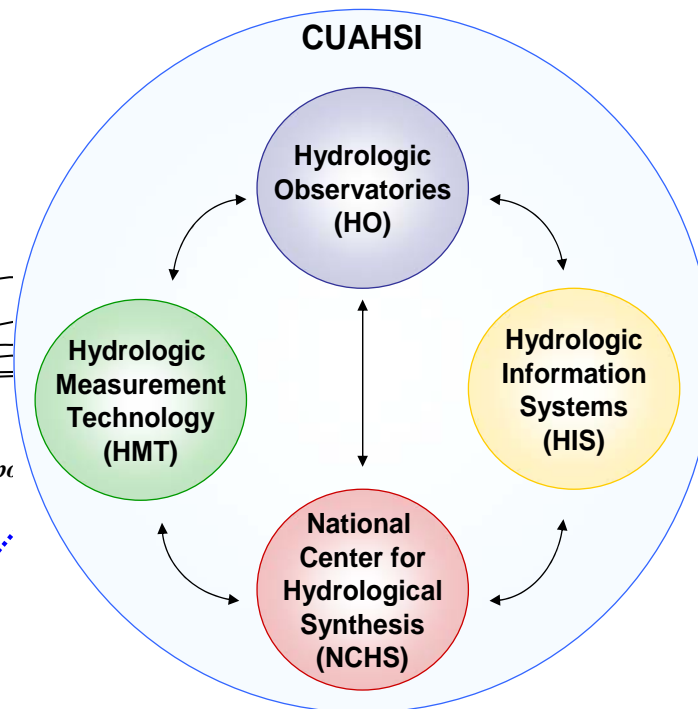


# Hydrology Synthesis – CUAHSI/NSF



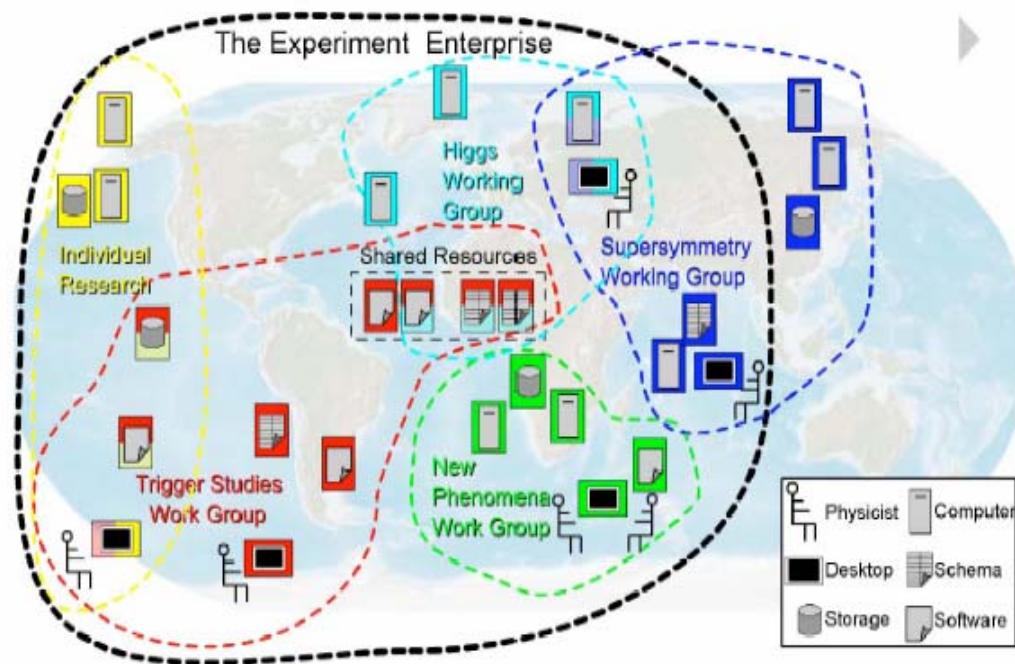
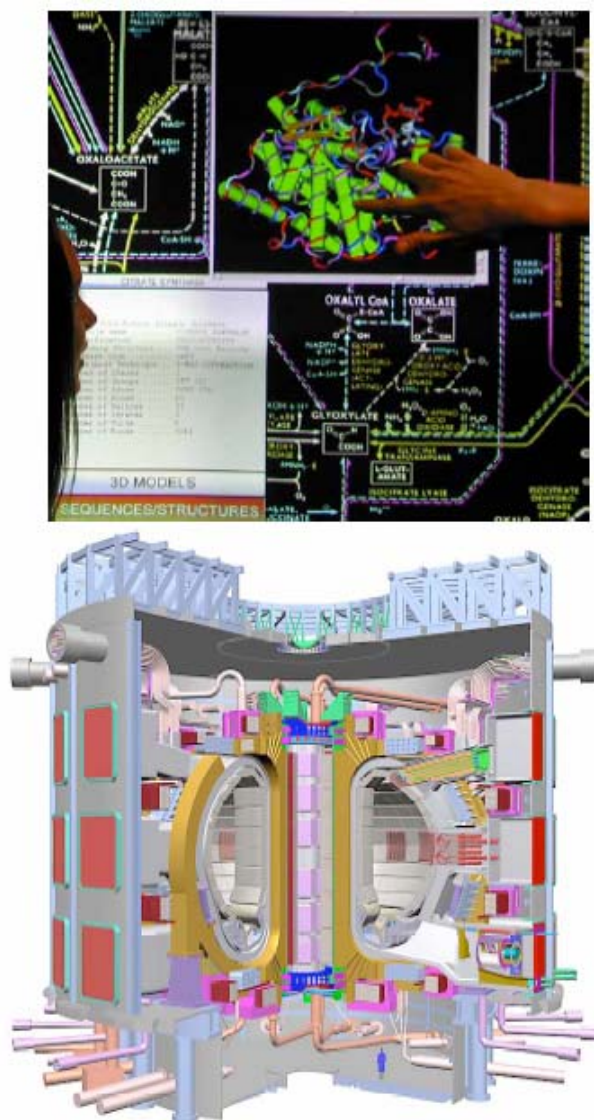
## HydroView

### CUAHSI





# Science Has Become a Team Sport



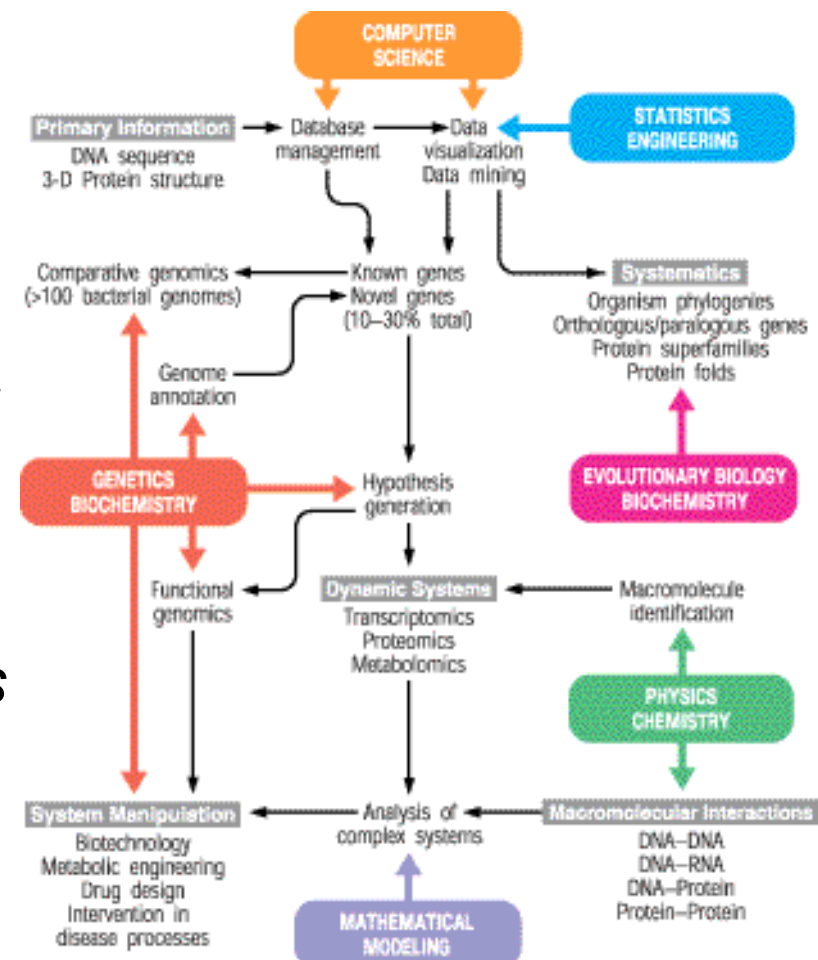
from Dave Schissel, GA



# Teams Sharing Data and Expertise



**Systems Biology:** “studying *biological systems by systematically perturbing them (biologically, genetically or chemically); monitoring the gene, protein, and informational pathway responses; integrating these data; and ultimately formulating mathematical models that describe the structure of the system and its responses to individual perturbations*” (Ideker et al., 2001 Annu. Rev. Genom. Hum. Genet. 2:343)



Konopka, 2004 *ASM News* 70:163



# Emerging Collaboration Modes



- Synchronous/Meetings
  - Seminar
  - Tutorial
  - Presentation
  - Working group
  - Brainstorming
- Asynchronous/Shared work products
  - Document editing
  - Code development
  - Shared data
  - Web portal
  - Workflow
- Semi-synchronous
  - Chat/presence
  - Shared web spaces - wikis
  - E-mail, blogs, etc





# Access Grid



- Original design
  - Support group-to-group interaction
  - Enable spontaneous meetings
  - Feeling of being down-the hall
  - Well designed and debugged space
  - Immersive audio and display
  - Based on existing tools
  - Dramatically improved launching and coordination
  - Natural interaction between participants
- Actual usage
  - Supports group-to-group interaction
  - Meetings are scheduled and tested
  - Debugging of settings and technologies often required
  - Immersive audio good video
- Community commitment and tolerance has been remarkably high



# Issues



- Multicast has not been as reliable as had been hoped
- Difficult to pursue research and deployment with limited resources
- Robustness of the software has been an issue
- Community contributions to the code have been limited
- Ability to adopt external developments has been limited
- Documentation has not kept up with the latest versions and best practices
- Minimal tuning, feedback, and debugging tools
- Cybersecurity mechanisms have caused problems



# Software Robustness and Support



- Thorough testing before release
- People dedicated to support function
- Issue tracking
- Consistent documentation
- Interoperability
- Tuning procedures and installation testing
- Tutorials
- User group
- Steering committee
- Need to pursue an infrastructure program providing these capabilities – coordinated internationally



# Networking



- IP Multicast has not become reliable
- Cybersecurity issues will likely make it less reliable over time
- Peer-to-peer overlay networks have proven that reliable dissemination networks can be built
- Combination of multicast and peer-to-peer networking with continuous testing and reconfiguration is needed





# Cybersecurity



- Collaborative tools are developed independent of operational cybersecurity considerations
  - Implications of site mechanisms
  - Protections from malicious code
  - Vulnerability testing
  - Interoperability with site cybersecurity mechanisms
- Typically there is a long process of debugging prototype deployments
  - Negotiating ports and protocols with each site's cybersecurity group
  - Debugging unexpected behaviors
  - Debugging security mechanisms
  - Identifying causes of performance problems



# Threats



- Viruses
- Worms
- Malicious software downloads
- Spyware
- Stolen credentials
- Insider Threat
- Denial of service
- Root kits
- Session hijacking
- Agent hijacking
- Man-in-the-middle
- Network spoofing
- Back doors
- Exploitation of buffer overflows and other software flaws
- Phishing
- Audits / Policy / Compliance
- ?????



## Example - Credential Theft



- Widespread compromises
  - Over 20++ sites
  - Over 3000+ computers
  - Unknown # of accounts
  - Very similar to unresolved compromises from 2003
- Common Modus Operandi
  - Acquire legitimate username/password via keyboard sniffers and/or trojaned clients and servers
  - Log into system as legitimate user and do reconnaissance
  - Use “off the shelf” rootkits to acquire root
  - Install sniffers and compromise services, modify ssh-keys
  - Leverage data gathered to move to next system
- ***The largest compromises in recent memory (in terms of # hosts and sites)***



# Collaboration Tools and Cybersecurity



- Collaborative tools building trust relationships across sites
- Extensive cross-site communication
- Many ports and protocols employed – high bandwidth
- Encrypted connections
- Multicast traffic difficult to police
- Servers need to be able to allow incoming connections
- Need to design cybersecurity considerations and coordination into the next generation of collaborative tools





# Next Generation Capabilities



- Whiteboards
- Shared screens
- Shared visualizations
- Security
- Video improvements
- Audio improvements
- New collaboration paradigms!
- Need well defined methods for migrating these capabilities as they are developed into the system so they are maintained and available broadly



# Proposed Path Forward



- Infrastructure development effort
  - Improved robustness and testing
  - Maintained documentation
  - Help desk (not just volunteer)
  - Tools for improved self-testing
  - Issue tracking
  - Interoperability with other systems
  - Strong community involvement
  - Well defined contribution model
  - Peer-to-peer overlay network to automatically adapt to lack of multicast
  - Internationally coordinated
- Research development effort
  - Next generation capabilities
  - Community involvement
  - Next generation paradigms
  - Beta tests with leading edge communities
  - Strong ties to user communities